DOCUMENT RESUME

ED 399 297 TM 025 612

AUTHOR Johanson, George A.; Gips, Crystal J.
TITLE In Praise of the Free-Response Item.
PUB DATE Apr 91
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Likert Scales; *Mail Surveys; *Principals; Secondary Education; Teacher Characteristics; *Teacher Qualifications; *Test Construction
IDENTIFIERS *Free Response Test Items; Open Ended Questions; Paired Comparisons

ABSTRACT

A mailed national survey of 500 randomly selected secondary school principals, which yielded 271 responses, used a single free-response item, 10 Likert-type items, and 36 paired-comparison items. The study was an attempt to determine the qualities that principals look for in selecting teacher candidates and whether there was a bias against the most cognitively able candidates. The single free-response item, "What is the single most important quality you look for in a teaching candidate?," was most valuable in clarifying responses and priorities among the qualities considered. The free response item enabled the researcher to validate the choice of attributes being rated and identified some assumptions of respondents that might not have been apparent otherwise. It provided additional insight and allowed the researcher to investigate the idea that respondents' answers in the highly structured format may be quite different from those in an open-ended or free-response format. (Contains two tables and seven references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
GEORGE JOHANSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

In Praise of the Free-Response Item

A paper presented at the annual meeting of the
American Educational Research Association
Chicago, IL
April 1991

ED 399 297

George A. Johanson & Crystal J. Gips
College of Education
Ohio University

Objectives

A mailed national survey of 500 randomly selected secondary school principals (Johanson & Gips, 1989) yielded 271 responses and utilized a single free-response item, 10 Likert items, and 36 paired-comparison items. The study was an effort to: 1. prioritize the qualities that secondary school principals look for in selecting a teacher candidate and 2. see if there was a bias against the most cognitively able--the best and brightest--candidates. The use of the differing item formats allowed for a better understanding of the preferences; the single free-response item proved most valuable. More specifically, the results of the three item formats suggest relationships between respondents' choices in hypothetical situations on a research instrument and real choices in the situation being studied.

TM025612

BEST COPY AVAILABLE

A literature review, two pilot studies, and extended discussions identified a list of 9 qualities to be investigated: rapport with students (R), communication/instructional skills (CIS), concern and caring (C&C), enthusiasm (E), cooperative/flexible attitude (CF), dedication to the profession (D), integrity and character (I&C), intellectual capacity (IC), and subject-area knowledge (SAK). The 10 Likert items were based on each of the above qualities, but CF was divided into "the ability to work cooperatively within our educational structure" (COOP/STR), and "a cooperative and flexible attitude toward students and staff", (COOP/SS). The purpose was to note distinctions in the meaning of cooperation (significant differences were not found).

Methods and Results

The scale values for importance ratings of the qualities using the different item formats appear in Table 1. The paired-comparisons were scaled using both

---

insert Table 1 about here

---

one-dimensional non-metric scaling (Kruskal & Wish, 1978) and Thurstone's Case V method (Dunn-Rankin,

1983). The results were nearly identical (r=0.99), and only the Thurstone scale values are presented. The zero value on the Thurstone scaling is arbitrary. The Likert items were scored 1-5 for responses from strongly disagree to strongly agree, respectively. The initial impression is that the Likert and paired-comparison formats tend to be in more agreement with each other than with the responses to the free-response format item: "What is the single most important quality you look for in a teaching candidate?". See Table 2.

---

insert Table 2 about here

The 258 free-responses were classified by both authors according to the nine qualities under consideration. The choice of the previously identified attributes was validated when 196 (76%) of the free-responses were found to be within existing categories. The purposely omitted "experience" (N=24), the "too-general" (N=7), and the "pejorative" (N=19) response classifications accounted for the majority of the additional answers. The low rating from the free-response item of the most

important quality from the paired-comparisons, I&C, may well be due to its being assumed by the majority of respondents in the same way that many other qualities, such as drug-dependence, went unmentioned in the free-response item but would surely have been seen as important if offered in the Likert format. If I&C is removed from the group, the correlations between the scales are much improved (see Table 2). The interpretation of the importance rating from either the Likert or paired-comparisons for I&C is much different in light of the free-response item: I&C is indeed very important when offered, that is, when recognition is possible, but may very well not be considered when not offered as an option and is therefore dependent upon recollection. Without a free-response item, such clarifications would be impossible and conclusions could be misleading and overly format-dependent. The improved correlations among the free-response, Likert, and paired scalings confirmed the free-response item's implication that I&C might best be omitted.

Each item format has rather specific advantages and liabilities. One benefit derived from the paired-comparison format is that the items are forced-choice

and the desire on the part of some respondents to see all of the qualities as highly desirable can not be satisfied. That is, forced-choices are relative or norm referenced while Likert items are absolute or domain referenced. Even with the very strong wording used in the Likert items ("...would tend to eliminate from further consideration any teaching candidate who failed to demonstrate..."), this scale suffered some ceiling effect with an overall mean response of 4.0. However, the benefits of a forced-choice become a liability if the choice in reality is not forced. An item-type can become artificial in such circumstances (Alwin & Krosnick, 1985). Choosing between a forced and unforced format is problematic when there is little theory to guide the researcher. The free-response item is admittedly difficult to code (Baldwin et. al., 1988), but may permit clarifications to either or both of the former item types.

An issue that needs clarification is the appropriateness of a one-dimensional scaling of the identified qualities. In fact, a PCA of the 10 Likert items revealed a two or three-dimensional solution (affective, cognitive, and possibly cooperative) and a

two dimensional non-metric MDS of the paired comparisons yielded a better fit than the one-dimensional scaling. In addition, the fit of the Thurstone Case V scaling was poor. As recommended by Guilford (1954), a Case III scaling was undertaken to determine if the cause of the misfit could be due to unequal discriminental dispersions. This appeared not to be the case, as the Case III scaling did not provide a substantially better fit to the data and the correlation with the Case V scale values was 0.99. Another possible source of difficulty was I&C. This was removed from the pairs and the remaining 8 characteristics were rescaled with the Case V methodology. As expected, the Case III scaling and the reduced Case V scaling correlated 0.99 with the original Case V values for the remaining 8 qualities.

The misfit of the one-dimensional model was thus likely due to the multi-dimensionality of the data. However, to compare the principals' perceptions of the relative importance of cognitive qualities and affective qualities, we must put them on a common scale in much the same way that we may compare apples to oranges in a supermarket. The relevant question

regarding the validity of the scale values is whether the various methods yield scale values that converge and, hence, give evidence that the scale values are reasonably independent of the methodology.

A related issue is the use of proportions to scale the free-response items. Micceri (1990) recommends the use of a logarithmic transformation of the proportions. With the present data, the transformed proportions correlated 0.99 with the untransformed proportions and the transformation was deemed unnecessary.

As for the question of bias against the most cognitively able, the survey had a resume of a candidate with both GPA and college attended manipulated, each with two levels. There was no interaction present for the entire sample and both main effects were statistically significant in a direction that indicated a preference for the "best and brightest". The main effects were not unexpected, but we had expected to find a significant interaction: the most desirable candidate coming from the less prestigious college but having the highest GPA. Evidence of such an interaction was lacking within any identifiable subsample until the free-response item was

used.

When only the principals indicating either of the cognitive qualities (SAK or IC) as their most important quality on the free-response item were considered, it was noted that the proportion of students receiving free or reduced lunch (FRL) was significantly correlated with the resume rating. This was not the case for the entire sample. A two-way analysis of covariance, controlling for the effect of FRL, indicated the hypothesized interaction. It might seem unexpected that the interaction would tend to exist only in those principal's that would first identify a cognitive trait as most desirable. However, the overall low ratings for the cognitive qualities might, in fact, indicate that only those principals for whom the cognitive qualities are important (as indicated by the free-response item) would demonstrate the hypothesized interaction.

One forced-choice item was designed to separate the principals with respect to cognitive and affective priorities. The item asked if, all else being equal, the principal would prefer to hire an experienced teacher without subject-matter knowledge or a non-

experienced candidate with subject-matter knowledge. Surprisingly, even the group of principals who identified subject-matter knowledge on the free-response item tended to prefer the experienced teacher. Needless to say, this item did not serve to define subgroups in the way intended and failed to define the subpopulation in which the interaction was present; only the free-response item identified the subgroup of interest.

Conclusions

In general, it may be concluded that the information obtained using a survey may depend heavily on the format of the questions asked and that the free-response format might well be a worthy addition to any survey, but especially to one that is attempting to scale objects or attributes. In particular, we would conclude that the use of a free-response item

1. enables the researcher to validate the choice of attributes being rated, helping to identify both omissions and extraneous items/attributes
2. identifies assumptions of the respondents that may have been difficult or impossible to anticipate

3. can give rather unique insight, perhaps enabling the respondents to be grouped in ways that are not possible without the free-response format

4. allows the researcher to investigate the notion that respondents' answers in a highly structured format are sometimes quite different from those in an open-ended or free-response format. That is, the use of a free-response item attempts to respond to the problem that a survey's structure influences a respondent's reply in a way that may be inconsistent with their unconstrained thoughts and perhaps even their actions

5. is economical to use.

In fact, we might conclude that the free-response item is a "best buy" for much survey research.

References

Alwin, D. F., & Krosnick, J. A. (1985). The measurement of values in surveys: A comparison of ratings and rankings. Public Opinion Quarterly, 49(4), 535-552.

Baldwin, C. L., Collins, J. R., Kostenbauer, T., & Murphy, C. B. (1988). Research choices for measuring outcome of high school groups. Journal for Specialists in Group Work, 13, 2-8.

Dunn-Rankin, P. (1983). Scaling methods. Hillsdale, NJ: Lawrence Erlbaum Associates.

Guilford, J. P. (1954). Psychometric Methods. New York: McGraw Hill.

Johanson, G. A., & Gips, C. J. (1989, April). Secondary school hiring practices: An exploratory study. Paper presented at the meeting of the American Educational Research Association, Boston, MA.

Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling. Beverly Hills: Sage Publications.

Micceri, T. (1990). Proportions, pitfalls and pendulums. Educational and Psychological Measurement, 50, 769-774.